CAAP Quarterly Report (Oct-Dec 2018)

Date of Report:	April 15, 2019
Prepared for:	U.S. DOT Pipeline and Hazardous Materials Safety Administration
Contract Number:	Cooperative Agreement #693JK31850011CAAP
Project Title:	Development of a prediction model for pipeline failure probability based on learning from past incidents and pipeline specific data using artificial neural network (ANN)
Prepared by:	Texas A&M Engineering Experiment Station
Contact Information:	James C. Holste j-holste@tamu.edu (979)845-3384
For quarterly period ending:	March 31, 2019

Business and Activity Section

(a) Contract Activity

The contract activities have been started.

(b) Status Update of Past Quarter Activities

Dr. Noor Quddus (postdoctoral research associate) is leading the team for the project. Graduate students worked in this quarter are Syeda Zohra Halim (graduating Ph.D. student), Mengxi Yu (4th year Ph.D. candidate), and Harold Escobar (3rd year PhD student). The team started examining incident data available at PHMSA website.

(c) Cost share activity

Due to some administrative challenges, no costs have been charged during this quarter. We will adjust the expenses in the coming quarters.

(d) Task 1: Development of methodology for creating root cause analysis reports

Only task-1 has been scheduled for this quarter.

Detailed discussion and descriptions:

1. Background and Objectives in the 2nd Quarter

In order to develop the ANN for building a predictive model, data of what contributed to past incidents will have to be gathered. All incidents will be investigated to determine what went wrong. Different root cause analyses will usually identify different causes and investigation reports may express these causes in a variety of ways though certain management/organization related causes may be common in several incidents. Thus, for the selection of inputs, it would be essential to determine a reference that defines what will actually be termed as root cause. Using taxonomy so that similar terms are used to refer to related root causes can help tackle this issue. The first challenge would be to build a set of cue words or taxonomy so that root causes analysis conducted for different incidents identify similar causes using similar terms and these causes will have to be identified in terms of measurable deviations/indicators so that they can later on be compared with deviations existing in a system to understand if the system is reaching an unsafe state. If a set of cue words are developed to produce all reports, extraction of information using automated systems based on text mining or data mining can be used. Task 1 will focus on determining how root cause analysis can be reported so that the information can be applied for prediction based on current condition.

2. Analysis in the 1st Quarter

Incident data for different pipeline systems has been downloaded from the PHSMA website. Summary of the data is shown in Figure 1. Total number of incidents in the hazardous liquid and CO2 pipeline system is the highest for a relatively lower mileage. This dataset involves wide varieties of chemicals involved. Because of these two reasons, the hazardous liquid and CO2 pipeline system has been selected for further examination.

Figure 1: Incident data for different pipeline systems [1]

Pipeline system	**Start year**	**Mileage (as of 2017) [2]**	**Total incidents (as of 2019) [3]**
Hazardous liquid and CO2	1986	215,628	9,690
Gas distribution	1986	2,223,209	4,301
Natural Gas transmissions and gathering	1986	318,807	3,422
LNG facilities and storage	2011	54,146,831 (capacity in bbl)	13

The incident reporting system has been changed over the years [4]. The number of reported data fields has been increased and after careful observation, it was found that the Hazardous liquid and CO2 flagged files has the following reported fields [5]:

1986-2011: 68 fields (columns)
2002-2009: 237 fields (columns)
2010 – present: 606 fields (columns)

For the sake of consistency in analysis, we have selected an even smaller subset of data 2010-present, which has 606 data fields and as of 1/31/2019, 3600 incident records [6]. The sub-fields (2010 – Present) are looked through carefully together with the corresponding incident reporting form. Since there is a tremendous amount of information in the file, fields in the file are grouped to summarize the data for the purpose of understanding the database. According the incident reporting form, the causes reported are as apparent causes and sub-causes. Depending on apparent causes, contributing factors or root cause may be required to be reported. Example: "List contributing factors" under "G5-Material Failure of Pipe or Weld", "Additional factors that contributed to the equipment failure", "CGF-DIRT Root Cause) under G3-Excavation Damage".

The pipeline incident investigation reports were also downloaded and examined [7]. A detail and systematic study will be conducted in the next quarter.

Pipeline incident reporting system and incident data from Canada National Energy Board [8] and European Gas Pipeline Incident Data Group (EGIG) [9] have been examined. A detail and systematic study will be conducted in the next quarter.

3. Results and Discussions

Initial screening has been done for the following three areas:

1. Pipeline incident data for Hazardous liquid and CO2, gas distribution, natural gas transmission, and LNG facilities and storage from PHMSA.
2. Incident data from Canada National Energy Board (NEB) and European Gas Pipeline Incident Data Group (EGIG).
3. Pipeline incident investigation reports from PHMSA and NTSB.

Incident reporting systems and root cause identification and categorization methods implemented by PHMSA, NEB and (EGIG) were examined.

Pipeline incident data for Hazardous liquid and CO2 (2010 – Present) has being studied in detail.

4. Future work

Task 1 will be continued in the next quarter. Pipeline incident data for Hazardous liquid and CO2 (2010 – Present) dataset will be examined for categorization of causes and sub-causes. Cause classification by NEB and EGIG will be studied and compared with PHMSA cause classification. Incident description will be studied carefully to identify potential new causes and refinement of old causes. Potential application of natural language processing (NLP) will be examined. Incident investigation reports from PHMSA and NTSB will be studied to identify new causes and extract insight relevant to root cause identification.

References

[1] https://www.phmsa.dot.gov/data-and-statistics/pipeline/source-data

[2] https://cms.phmsa.dot.gov/data-and-statistics/pipeline/annual-report-mileage-summary-statistics

[3] https://www.phmsa.dot.gov/data-and-statistics/pipeline/source-data)

[4] https://www.phmsa.dot.gov/data-and-statistics/pipeline/history-phmsa-incident-reporting-criteria
[5] https://www.pipeline incident data\flagged data\downloaded data\hazardous liquid
[6] https://www.pipiline incident data\flagged data\downloaded data\hazardous liquid\ hl2010toPresent
[7] https://www.phmsa.dot.gov/safety-reports/pipeline-failure-investigation-reports
[8] https://www.neb-one.gc.ca/sftnvrnmnt/sft/dshbrd/mp/dt-eng.html
[9] https://www.egig.eu/reports